

PCT

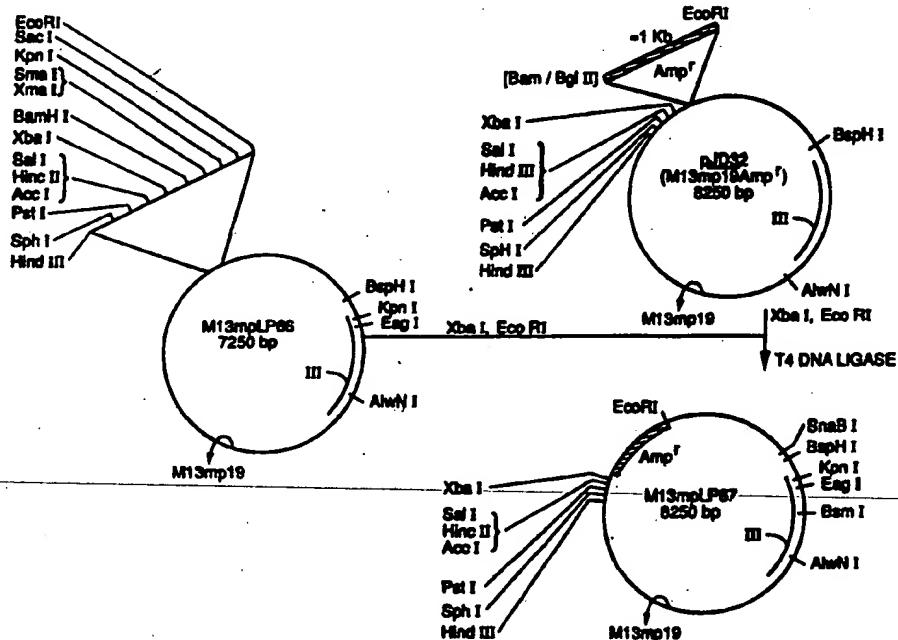
WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :  C12N 15/10, 15/62, 15/34 C12N 15/70, 15/00 // C07K 7/06		A1	(11) International Publication Number: WO 91/18980  (43) International Publication Date: 12 December 1991 (12.12.91)
(21) International Application Number: PCT/US91/03332			(81) Designated States: AT (European patent), AU, BE (European patent), CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FR (European patent), GB (European patent), GR (European patent), IT (European patent), JP, LU (European patent), NL (European patent), SE (European patent).
(22) International Filing Date: 13 May 1991 (13.05.91)			
(30) Priority data: 533,180 1 June 1990 (01.06.90) US			
(71) Applicant: CETUS CORPORATION [US/US]; 1400 Fifty-Third Street, Emeryville, CA 94608 (US).			Published <i>With international search report.</i>
(72) Inventor: DEVLIN, James, J. ; 1146 Upper Happy Valley, Lafayette, CA 94549 (US).			
(74) Agent: GIOTTA, Gregory, J.; Cetus Corporation, 1400 Fifty-Third Street, Emeryville, CA 94608 (US).			

(54) Title: COMPOSITIONS AND METHODS FOR IDENTIFYING BIOLOGICALLY ACTIVE MOLECULES



(57) Abstract

Methods and compositions are described for producing a random peptide library for identifying molecules that have biological activity, and that would have applications in the clinical setting as prophylactics and/or therapeutics.

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MG	Madagascar
AU	Australia	FI	Fieland	ML	Mall
BB	Barbados	FR	France	MN	Mongolia
BE	Belgium	GA	Gabon	MR	Mauritania
BF	Burkina Faso	GB	United Kingdom	MW	Malawi
BG	Bulgaria	GN	Guinca	NL	Netherlands
BJ	Benin	GR	Greece	NO	Norway
BR	Brazil	HU	Hungary	PL	Poland
CA	Canada	IT	Italy	RO	Romania
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland	KR	Republik of Korca	SN	Senegal
CI	Côte d'Ivoire	LI	Liechtenstein	SU	Soviet Union
CM	Cameroon	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TG	Togo
DE	Germany	MC	Monaco	US	United States of America

## COMPOSITIONS AND METHODS FOR IDENTIFYING BIOLOGICALLY ACTIVE MOLECULES

5 This invention is in the area of molecular biology and presents compositions and methods for identifying biologically active molecules that will be used as medicaments in the clinical setting. The invention also has applications in the fields of clinical immunology and pharmacology.

10 A primary goal of molecular biology is to identify biologically active molecules that have practical clinical utility. The general approach taken by molecular biologists has been to initially identify a biological activity of interest, and then purify the activity to homogeneity. Next, assuming the molecule is a protein, the protein is sequenced and the sequence information used to generate synthetic DNA oligonucleotides that represent potential codon combinations that encode the protein of interest. The 15 oligonucleotide is then used as a probe to probe a cDNA library derived from messenger RNA that was in turn derived from a biological source that produced the protein. The cDNA sequence so identified may be manipulated and expressed in a suitable expression system.

20 A second, more recent approach, termed expression cloning, avoids purifying and sequencing the protein of interest, as well as generating oligonucleotide probes to screen a cDNA library. Rather this procedure consists of initially ascertaining the presence of a biologically active molecule, generating cDNA from messenger RNA and directly cloning the cDNA into a suitable expression vector. The vector is typically an expression plasmid that is transfected or micro-injected into a suitable host cell to realize 25 expression of the protein. Pools of the plasmid are assayed for bioactivity, and by narrowing the size of the pool that exhibits activity, ultimately a single clone that expresses the protein of interest is isolated.

30 Aside from the above approaches, it is, of course, well known that bioactive molecules other than proteins are constantly being isolated and screened in large numbers using traditional screening regimens well known to those that work in this field. Additionally, after a drug is identified and its chemical structure elucidated, attempts are made to synthesize more active versions of the drug by rational drug design.

35 Previously, it was suggested that an "epitope library" might be made by cloning synthetic DNA that encodes random peptides into filamentous phage vectors. Parmley and Smith, 1988, *Gene*, 73:305. It was proposed that the synthetic DNA be cloned into the coat protein gene III because of the likelihood of the encoded peptide becoming part of pIII without significantly interfering with pIII's function. It is known

that the amino terminal half of pIII binds to the F pilus during infection of the phage into *E. coli*. It was suggested that such phage that carry and express random peptides on their cell surface as part of pIII may provide a way of identifying the epitopes recognized by antibodies, particularly using antibody to affect the purification of phage  
5 from the library. Pamley and Smith, 1988, *Gene*, 23:305. Unfortunately, this approach to date has not produce any useful biologically active molecules.

Previous investigators have shown that the outer membrane protein, LamB, of *E. coli* can be altered by genetic insertion to produce hybrid proteins having inserts up to about 60 amino acid residues. Charbit, A., Molla A., Saurin, W., and Hofnung,  
10 1988, *Gene*, 70:181. The authors suggests that such constructs may be used to produce live bacterial vaccines. See also, Charbit, A., Boulain, J. C. Ryter, A. and Hofnung, M., 1986, *EMBO J.*, 5(11):3029; and, Charbit, A., Sobczak, E., Michael, M.L., Molla, A., Tiollais, P., and Hofnung, M., 1987, *J. Immunol.*, 139(5):1658.

The procedures that are presently used to identify protein bioactive molecules,  
15 as well as small molecular weight molecules, require a significant commitment of resources which often limit the progress of such projects. Thus, other methods that facilitate the identification of bioactive molecules are keenly sought after, and would have wide applicability in identifying medicaments of significant practical utility.

One aspect of the invention is the description of a method for constructing a  
20 library of random peptides that may be used to identify bioactive peptides.

A second aspect of the invention is the description of a method and compositions for generating a random peptide library by cloning synthetic DNA, which encodes the random peptide sequences, into an appropriate expression vector.

A third aspect of the invention is the description of a random peptide library  
25 wherein the peptides are expressed on the surface of infectious filamentous phage that permits screening greater than  $10^6$  random peptide sequences at a time.

A fourth aspect of the invention is the identification of biotin inhibitable streptavidin binding peptides using a random peptide library of about  $2 \times 10^7$  different fifteen residue peptides. The peptides have a dipeptide amino acid consensus sequence of His-Pro, and may have a tripeptide amino acid consensus sequence of His-Pro-Gln.  
30

A fifth aspect of the invention is a description of a defined filamentous phage expression system having random peptides expressed as a fusion protein with an appropriate phage protein such that the random peptides do not substantially interfere with the normal biological activities of the phage.

35 A sixth aspect of the invention is a description of a defined filamentous phage expression system having random peptides expressed as a fusion protein with an appropriate phage protein, the phage protein being expressed on the surface of the

phage such that the random peptides are thus exposed and readily available for screening to determine their biological properties.

A seventh aspect of the invention is a description of a preferred filamentous phage expression system having random peptides expressed as a fusion protein with an appropriate phage protein, the phage protein being expressed on the surface of the phage, and the random peptides positioned at the amino terminal region of the phage protein. Positioning the random peptides at the amino terminal region facilitates screening the peptides for biological activity.

An eighth aspect of the invention is a description of a preferred filamentous phage expression system having random peptides expressed as a fusion protein with an appropriate phage protein, the phage protein being expressed on the surface of the phage, and the random peptides positioned at the amino terminal region of the phage protein, thereby facilitating screening the peptides for biological activity, wherein the fusion protein has the following construct:



V stands for one or more amino acids of the wild type phage protein found at the amino terminus; RP stands for the random peptide sequence; L stands for a spacer sequence that facilitates presentation of the random peptide sequence to screening reagents; and  $V_1$  stands for the amino acid sequence of the phage protein.

In addition to the foregoing, other aspects of the invention will become apparent upon reading the detailed description of the invention presented below.

Figures 1 and 2 show the construction of M13LP67.

Table 1 shows the results of enriching for M13LP67 virions that encode streptavidin binding peptides that are present on the surface of the virions. The Table also shows that biotin inhibits the binding of certain of the peptides to streptavidin.

Table 2 presents the predicted amino acid sequence of the random peptides encoded by several streptavidin binding phage.

The invention described herein draws on previously published work and pending patent applications. By way of example, such work consists of scientific papers, patents or pending patent applications. All of these publications and applications, cited previously or below are hereby incorporated by reference.

Described herein is a method for producing a library consisting of random peptide sequences. The library can be used for many purposes including identifying and selecting peptides that have a particular bioactivity, as well using such peptides, in those instances where they have ligand properties, to isolate and purify ligand binding molecules. An example of a ligand binding molecule would be a soluble or insoluble

cellular receptor (i.e. membrane bound receptor), but would extend to virtually any molecule, including enzymes, that have the sought after binding activity.

5 "Cells" or "recombinant host" or "host cells" are often used interchangeably as will be clear from the context. These terms include the immediate subject cell, and, of course, the progeny thereof. It is understood that not all progeny are exactly identical to the parental cell, due to chance mutations or differences in environment.

10 As used herein the term "transformed" in describing host cell cultures denotes a cell that has been genetically engineered to produce a heterologous protein that possesses the activity of the native protein. Examples of transformed cells are described in the examples of this application. Bacteria are preferred microorganisms for producing the protein. Synthetic protein may also be made by suitable transformed yeast and mammalian host cells.

15 "Operably linked" refers to juxtaposition such that the normal function of the components can be performed. Thus, a coding sequence "operably linked" to control sequences refers to a configuration wherein the coding sequence can be expressed under the control of these sequences.

20 "Control sequences" refers to DNA sequences necessary for the expression of an operably linked coding sequence in a particular host organism. The control sequences which are suitable for prokaryotes, for example, include a promoter, optionally an operator sequence, a ribosome binding site, and possibly, other as yet poorly understood, sequences. Eucaryotic cells are known to utilize promoters, polyadenylation signals, and enhancers.

25 "Expression system" refers to DNA sequences containing a desired coding sequence and control sequences in operable linkage, so that hosts transformed with these sequences are capable of producing the encoded proteins. In order to effect transformation, the expression system may be included on a vector; however, the relevant DNA may then also be integrated into the host chromosome.

30 The term "mature" protein is known in the art, and is intended to denote proteins that have a peptide segment of the protein removed during in vivo, or in vitro synthesis.

One embodiment of the invention is the description of methods and compositions for constructing a library of random peptides consisting of cloning synthetic DNA, which has a degenerate coding sequence that encodes the random peptide sequences, into an appropriate expression vector that has control sequences in operable linkage for regulation and expression of the random peptide sequence. The vector is then inserted into an appropriate host cell compatible with expression of the random peptide sequence. For example, depending on the nature of the vector, it may

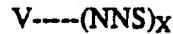
be transfected, transformed, electroporated, or in other ways inserted into the host cell. The random peptide sequence may be expressed as part of a fusion protein construct. Subsequently, cells that harbor the random peptide sequences are selected, isolated, and grown up. The random peptide sequence may be expressed in soluble or insoluble form (i.e. membrane bound), and, if desired, purified prior to being employed in biological screens aimed at identifying its properties.

5 The preferred embodiment synthetic DNA degenerate coding sequence that encodes the random peptide is exemplified by the following formula:



10 N is either an equal mixture of the deoxynucleotides A, G, C and T, or a mixture consisting of equal amounts of A, T, C, and an excess of G, about 30% more than A. S is an equal mixture of C and G. X is the number of amino acid residues that make up the random peptide sequence, and will vary depending on the characteristics of the random peptide that is sought to be identified. Preferably X is greater than 6 but  
15 less than 16. However, it will be appreciated that while this is the preferred size of the random peptide, larger peptides are clearly intended to come within the scope of the invention, as described in detail below.

If the random peptide is constructed as part of a fusion protein, a preferred construct that contains the synthetic DNA degenerate coding sequence that encodes the  
20 random peptide may be exemplified by the following formula:



V stands for nucleotide sequences that encode amino acid (s) found at the amino terminus of a mature protein to which the degenerate coding sequence is fused; N is either an equal mixture of the deoxynucleotides A, G, C and T, or an equal  
25 mixture of A, T, C, and an excess of G, about 30%. S is an equal mixture of C and G. X is the number of amino acid residues that make up the random peptide sequence, and will vary depending on the characteristics of the random peptide that is sought to be identified. Preferably X is greater than 6, but less than 16.

It is important to note, and it will be apparent to those skilled in this art, that the  
30 size of the random peptide sequence, X, will vary depending on many parameters. A key exemplary parameter is the nature of the protein that is the fusion target. If such proteins have biological activities that contribute to the formation of the random peptide library, then determinative of the size of the random peptide insert will be whether it has deleterious biological effects. For example, it is known that the outer membrane  
35 protein, LamB, of *E. coli* can be altered by genetic insertion to produce hybrid proteins having inserts up to about 60 amino acid residues. Charbit, A., Molla A., Saurin, W.,

and Hofnung, 1988, Gene, 70:181. Above 60 residues LamB loses significant biological activity.

A more preferred embodiment random peptide cassette construct may be expressed as part of a fusion protein that has the following formula:



V, and (NNX)<sub>X</sub>, are as defined above, L stands for an amino acid spacer sequence that facilitates preservation of the active conformation of the random peptide, and serves to present the random peptide sequence to screening reagents. V<sub>1</sub> is the coding sequence of the protein to which the construct is fused.

10                  The synthetic DNA degenerate coding sequence is prepared using known oligonucleotide synthesis techniques. For instance, synthetic oligonucleotides may be prepared by the triester method of Matteucci et al., 1981, J Am Chem Soc, 103:3185 or using commercially available automated oligonucleotide synthesizers.

15                  It is worth noting that in a preferred embodiment of the invention, the constructs described above could be designed to encode a cysteine which would facilitate conjugating the peptide to an appropriate substrate. This in turn would facilitate assaying the biological properties of the peptides in those assay formats where it is advantageous to affix the peptide to a solid surface.

20                  A variety of vectors may be used to clone and express the random peptide sequence. Viral or plasmid vectors are preferred. Synthetic DNA that encodes the random peptides is most preferably cloned into viral vectors, and more preferably into bacteriophage vectors. Should the random peptide be expressed as a fusion protein, the preferred viral vectors are the λ gt series, preferably gt11 or derivatives thereof including λ gt11 Sfi-Not (Promega), and filamentous bacteriophage vectors, preferably M13, f1 and fd. Most preferred is M13 or derivatives thereof.

25                  A random peptide expressed as part of a fusion protein in λ gt11 is preferably realized by cloning the synthetic DNA into restriction sites at the region of the gene encoding the carboxyl terminal end of an appropriate target protein such as β-galactosidase. The resulting fusion proteins are conveniently assayed using standard λ gt11 screening assays, preferably the methods as described by Davis, R. W., and Young, R. A., in U. S. Patent No. 4,788,135.

30                  A favored embodiment fusion protein construct consists of filamentous bacteriophage in which the random peptide sequences are cloned into phage surface proteins. This has the advantage of screening the random peptide sequence as part of the phage, which is particularly advantageous since the phage may be applied to affinity

matrices at over  $10^{13}$  phage /ml, and thus screened in large numbers. Secondly, since a particular phage population may be removed from an affinity matrix and maintain substantial infectious activity, the random peptide may be amplified by subsequent infection of a suitable bacterial host.

5 For several reasons the preferred phage surface protein is the minor coat protein encoded by gene III of filamentous phage. For instance, foreign antigen epitopes can be expressed in the middle of gene III without significantly disrupting the infectious properties of the phage. Parmley and Smith, 1988, Gene, 73:305. Moreover, five copies of the gene III protein per virion are expressed in E. coli, thus rendering multiple 10 copies available for detection in the screening assays described below.

A preferred embodiment vector of the instant invention that was used to construct the random peptide library is a derivative of M13mp19, termed M13LP67. The construction of this vector, as well as cloning the synthetic degenerate oligonucleotides that encode the random peptide, was carried out using commonly 15 employed techniques known to the skilled practitioner of molecular biology. The reader is particularly referred to Maniatis et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Laboratory, New York, (1982, and 1989, volumes 1 and 2). In addition, many of the materials and methods described herein are also exemplified in Methods & Enzymology, 153-155, Editor Ray 20 Wu/Lawrence Grossman, Academic Press, Inc.. Volume 153 covers methods related to vectors for cloning DNA and for the expression of cloned genes. Particularly note worthy is volume 154, which describes methods for cloning cDNA, identification of various cloned genes and mapping techniques useful to characterize the genes, chemical synthesis and analysis of oligodeoxynucleotides, mutagenesis, and protein engineering. 25 Finally, volume 155 presents the description of restriction enzymes, particularly those discovered in recent years, as well as methods for DNA sequence analysis. These references are hereby incorporated in their entirety, as well as are additional references described below. A general description of the salient methods and materials used is presented here for the convenience of the reader.

30 More specifically, construction of suitable vectors containing the desired random coding sequence employs standard ligation and restriction methods wherein isolated vectors, DNA sequences, or synthesized oligonucleotides are cleaved, tailored, and religated in the form desired.

Site specific DNA cleavage is performed by treating with suitable restriction 35 enzyme(s) under conditions which are generally understood in the art, and the particulars of which are specified by the manufacturer of these commercially available restriction enzymes. See, e.g., New England Biolabs, Product Catalog. In general,

about 1  $\mu$ g of plasmid or DNA sequence is cleaved by one unit of enzyme in about 20  $\mu$ l of buffer solution. In the examples herein, typically, an excess of restriction enzyme is used to insure complete digestion of the DNA substrate. Incubation times of about 1 to 2 hours at about 37°C are workable, although variations can be tolerated. After each 5 incubation, protein is removed by extraction with phenol/chloroform, and may be followed by ether extraction, and the nucleic acid recovered from aqueous fractions by precipitation with ethanol followed by chromatography using a Sephadex G-50 spin column. If desired, size separation of the cleaved fragments may be performed by polyacrylamide gel or agarose gel electrophoresis using standard techniques. A general 10 description of size separations is found in Methods in Enzymology, 1980, 65:499-560.

Restriction cleaved fragments may be blunt ended by treating with the large fragment of *E. coli* DNA polymerase I, that is, the Klenow fragment, in the presence of the four deoxynucleotide triphosphates (dNTPs) using incubation times of about 15 to 25 minutes at 20-25°C in 50 mM Tris pH 7.6, 50 mM NaCl, 6 mM MgCl<sub>2</sub>, 6 mM DTT 15 and 10 mM dNTPs. After treatment with Klenow, the mixture is extracted with phenol/chloroform and ethanol precipitated. Treatment under appropriate conditions with S1 nuclease results in hydrolysis of single-stranded portions.

Ligations are performed in 15-30  $\mu$ l volumes under the following standard 20 conditions and temperatures: 20 mM Tris-Cl pH 7.5, 10 mM MgCl<sub>2</sub>, 10 mM DTT, 33  $\mu$ g/ml BSA, 10 mM-50 mM NaCl, and 1 mM ATP, 0.3-0.6 (Weiss) units T4 DNA ligase at 14°C for "sticky end" ligation, or for "blunt end" ligations 1 mM ATP was used, and 0.3-0.6 (Weiss) units T4 ligase. Intermolecular "sticky end" ligations are usually performed at 33-100  $\mu$ g/ml total DNA concentration. In blunt end ligations, the total DNA concentration of the ends is about 1  $\mu$ M.

The vector construction employing "vector fragments", the vector fragment is 25 commonly treated with bacterial alkaline phosphatase (BAP) in order to remove the 5' phosphate and prevent religation of the vector. BAP digestions are conducted at pH 8 in approximately 150 mM Tris, in the presence of Na<sup>+</sup> and Mg<sup>+2</sup> using about 1 unit of BAP per  $\mu$ g of vector at 60°C for about 1 hour. Nucleic acid fragments are recovered 30 by extracting the preparation with phenol/chloroform, followed by ethanol precipitation. Alternatively, religation can be prevented in vectors which have been double digested by additional restriction enzyme digestion of the unwanted fragments.

For portions of the vector which have particular sequence modifications to 35 introduce a desired restriction site, site-specific primer directed mutagenesis was used. For example, it was desirable to modify M13mp19 to introduce restriction sites to produce M13LP67. Site-specific primer directed mutagenesis is now standard in the

art, and is conducted using a primer synthetic oligonucleotide complementary to a single stranded phage DNA to be mutagenized except for limited mismatching, representing the desired mutation. Briefly, the synthetic oligonucleotide is used as a primer to direct synthesis of a strand complementary to the phage, and the resulting 5 double-stranded DNA is transformed into a phage-supporting host bacterium. Cultures of the transformed bacteria are plated in top agar, permitting plaque formation from single cells which harbor the phage.

Theoretically, 50% of the new plaques will contain the phage having, as a 10 single strand, the mutated form; 50% will have the original sequence. The plaques are transferred to nitrocellulose filters and the "lifts" hybridized with kinased synthetic primer at a temperature which permits hybridization of an exact match, but at which the mismatches with the original strand are sufficient to prevent hybridization. Plaques which hybridize with the probe are then picked and cultured, and the DNA is recovered.

15 Details of site specific mutation procedures are described below in specific examples. However, site specific mutagenesis can be carried out using any number of procedures known in the art. These techniques are described by Smith, 1985, Annual Review of Genetics, 19:423, and modifications of some of the techniques are described in Methods in Enzymology, 154, part E, (eds.) Wu and Grossman (1987), chapters 20 17, 18, 19, and 20. A preferred procedure is a modification of the Gapped Duplex site-directed mutagenesis method. The general procedure is described by Kramer, et al., in chapter 17 of the Methods in Enzymology, above.

M13 may be propagated as either a virus, or a plasmid; and if propagated as a 25 plasmid, to facilitate identification of cells that harbor the plasmid, it is desirous that it carry a suitable marker gene. A representative marker gene is  $\beta$ -lactamase, and may be obtained from the plasmids pUC19 or pAc5 by polymerase chain reaction (PCR) amplification. pAc5 is described in WO 89/01029. It was amplified using appropriate primers and standard methods known in the art, or disclosed in U.S. Patent Nos. 4,683,195 issued July 28, 1987; 4,683,202 issued July 28, 1987; and 4,800,159 30 issued January 24, 1989 the latter of which is incorporated herein by reference in its entirety. A modification of this procedure involving the use of the heat stable Thermus aquaticus (Taq) DNA polymerase has been described and characterized in European Patent Publication No. 258017 published March 2, 1988 incorporated herein by reference in its entirety. PCR is conveniently carried out using the Thermal Cycler 35 instrument (Perkin-Elmer-Cetus) which has been described in European Patent Publication No. 236,069, published September 9, 1987 also incorporated herein by reference in its entirety.

In the constructions set forth below, correct ligations are confirmed by first transforming the appropriate *E. coli* strain with the ligation mixture. Successful transformants are selected by resistance to ampicillin, tetracycline or other antibiotics, or using other markers depending on the mode of plasmid construction, as is understood in the art. Miniprep DNA can be prepared from the transformants by the method of D. Ish-Howowicz et al., 1981, *Nucleic Acids Res.*, 9:2989 and analyzed by restriction and/or sequenced by the dideoxy method of F. Sanger et al., 1977, *Proc. Natl. Acad. Sci. (USA)*, 74:5463 as further described by Messing et al., 1981, *Nucleic Acids Res.*, 9:309, or by the method of Maxam et al., 1980, *Methods in Enzymology*, 65:499.

Host strains used to propagate M13 and derivatives include *E. coli* strains susceptible to phage infection, such as *E. coli* K12 strain DG98, and strain H249, which is a recA, supE, F', kan<sup>r</sup> derivative of MM294. H249 cells are favored if the cells are to be electroporated. The DG98 strain has been deposited with ATCC July 13, 1984 and has Accession No. 1965.

Depending on the host cell used, transformation is done using standard techniques appropriate to such cells. The preferred method is electroporation in a low conductivity solution as described by Dower, W.J., et al., 1988, *Nuc. Acids Res.*, 16:6127. Commercially available electroporation machines may be utilized, such as, for example those made by BTX. Other methods, however, may also be used. For example, the calcium treatment employing calcium chloride, as described by Cohen, S.N., et al., 1972, *Proc. Natl. Acad. Sci. (USA)* 69:2110, and modifications as described by Hanahan, D., 1983, *J. Mol. Biol.*, 166:557-580 are used for prokaryotes or other cells which contain substantial cell wall barriers. Several transfection techniques are available for mammalian cells without such cell walls. The calcium phosphate precipitation method of Graham and Van Der Eb, 1978, *Virology* 52:546 is one method. Transfection can be carried out using a modification (Wang et al., 1985, *Science* 228:149) of the calcium phosphate co-precipitation technique. Another transfection technique involves the use of DEAE-dextran (Sompayrac, L.M. et al., 1981, *Proc. Natl. Acad. Sci. USA* 78:7575-7578). Alternatively, Lipofection refers to a transfection method which uses a lipid matrix to transport plasmid DNA into the host cell. The lipid matrix referred to as Lipofectin Reagent is available from BRL. Lipofectin Reagent comprises an aqueous solution (deionized and sterile filtered water) containing 1 mg/ml of lipid (DOTMA:DOPE, 50:50). This liposome-mediated transfection is carried out essentially as described by Felgner, P.L., et al., 1987, *Proc. Natl. Acad. Sci. U.S.A.*, 84:7413. Lipofectin Reagent and DNA are separately diluted into serum free media so as to avoid gross aggregation which can occur when either

material is too concentrated. For example,  $0.5 \times 10^6$  cells are seeded onto a 60 mm tissue culture dish, and 1.5 ml of serum free media containing 1 to 20  $\mu\text{g}$  f DNA and a second solution of 1.5 ml serum free media containing about 30  $\mu\text{g}$  of Lipofectin are prepared. The diluted DNA and Lipofectin solutions are mixed and applied onto the 5 cells. Since transfection is inhibited by serum, the cells are washed well with serum free media before adding the Lipofectin/DNA mixture.

Depending on the properties of the particular random peptide sought, it may be assayed using one or more assay methods. Generally, those methods aimed at detecting the random peptide will detect the random peptide by its capacity to bind an 10 appropriate binding molecule. That is, a molecule that is sought to be tested for random peptide binding activity. Depending on the assay format, it may be labelled or unlabelled. Further, if it is desirable to do so, the random peptide sequence may be purified prior to being employed in the assay.

If the random sequence is part of a fusion protein, preferably a filamentous viral 15 surface protein, the presence of the random peptide sequence may be indicated by the binding of virus to a chosen binding molecule, and separating bound and unbound virus. In this way, virus that contains the random peptide of interest may be isolated, and subsequently amplified by infection of a suitable host cell. Confirmation that the 20 virus encodes a random sequence, as well as the predicted amino acid sequence, can be obtained using standard techniques, including the polymerase chain reaction, and DNA sequencing, respectively.

A random peptide sequence expressing virus may also be revealed using a 25 labelled binding molecule. As applied to viral fusion proteins, preferably lambda bacteriophage, one method is similar to a method described by Davis, R. W., and Young, R. A., in U. S. Patent No. 4,788,135, wherein replica plated phage plaques are contacted with a labelled binding molecule. Fusion proteins produced by phage that express the random peptide sequence will bind the labelled binding molecule, and the phage encoding them can then be identified, isolated and grown up from the replica plates.

30 Each of the above purification techniques may be repeated multiple times to enrich for the virus that encodes the random peptide of interest.

The binding molecule can be labelled with any type of label that allows for the detection of the binding molecule under the conditions employed, and preferably when 35 the random peptide sequence is bound to a support matrix. Generally, the label directly or indirectly results in a signal which is measurable and related to the amount of random peptide present in the sample. For example, directly measurable labels can include radio-labels (e.g.  $^{125}\text{I}$ ,  $^{35}\text{S}$ ,  $^{14}\text{C}$ , etc.). A preferred directly measurable label is an

enzyme, conjugated to the binding molecule, which produces a color reaction in the presence of an appropriate substrate (e.g. horseradish peroxidase/o-phenylenediamine). An example of an indirectly measurable label is a binding molecule that has been biotinylated. The presence of this label is measured by contacting it with a solution containing a labelled avidin complex, whereby the avidin becomes bound to the biotinylated binding molecules. The label associated with the avidin is then measured. A preferred example of an indirect label is the avidin/biotin system employing an enzyme conjugated to avidin, the enzyme producing a color reaction as described above. Other methods of detection are also usable. For example, avidin could be detected using an appropriately labelled avidin binding antibody.

The following examples are illustrative of various ways in which the invention may be practiced. However, it will be understood by those skilled in the art that the presentation of such examples, showing specific materials and methods, should not be construed as limiting the invention to what is shown in the examples, it being well known by the skilled practitioner that there are numerous substitutions that would perform similarly.

#### Example I

##### Degenerate Oligonucleotides Encoding Random Peptides

A degenerate oligonucleotide having the following structure was synthesized, and purified using methods known in the art:

5' CTTTCTATTCTCACTCCGCTGAA(NNS)<sub>15</sub>CCGCCTCCACCTCCACC 3'  
5'GGCCGGTGGAGGTGGAGGC GG(XXX)<sub>15</sub>TTCAGCGGAGTGAGAAT  
AGAAAGGTAC 3'

During the synthesis of (NNS)<sub>15</sub> mixture consisting of equal amounts of the deoxynucleotides A, C and T, and about 30% more G was used for N, and an equal mixture of C and G for S. X stands for deoxyinosine, and was used because of its capacity to base pair with each of the four bases A,G,C, and T. Reidhaar-Olson, J.F., and Sauer, R. T., 1988, Science, 24:53. Alternatively, other base analogs may be used as described by Habener, J., et al., 1988, PNAS, 85:1735.

Immediately preceding the nucleotide sequence that encodes the random peptide sequence is a nucleotide sequence that encodes alanine and glutamic residues. These amino acids were included because they correspond to the first two amino terminal residues of the wild type mature gene III protein of M13, and thus may facilitate producing the fusion protein produced as described below.

Immediately following the random peptide sequence, is a nucleotide sequence that encodes 6 prolines residues. Thus, the oligonucleotide encodes the following amino acid sequence:



5 Zzz denotes amino acids encoded by the degenerate DNA sequence.

As will be described below, the oligonucleotides were cloned into a derivative of M13 to produce a mature fusion protein having the above amino acid sequence, and additionally, following the proline residues the entire wild type mature gene III.

10

#### Example II

##### Construction of the Plasmid M13LP67

The plasmid M13LP67 was used to express the random peptide/ gene III fusion protein construct. M13LP67 was derived from M13 mp19 as shown in Figures 1 and 2.

15

Briefly, M13mp19 was altered in two ways. The first alteration consisted of inserting the marker gene,  $\beta$ -lactamase, into the polylinker region of the virion. This consisted of obtaining the gene by PCR amplification from the plasmid pAc5. The oligonucleotide primers that were annealed to the pAc5 template have the following sequence:

20

5' GCT GCC CGA GAG ATC TGT ATA TAT GAG TAA ACT TGG 3'  
5' GCA GGC TCG GGA ATT CCG GAA ATG TGC GCG GAA CCC 3'

Amplified copies of the  $\beta$ -lactamase gene were digested with the restriction enzymes BglII and EcoRI, and the replicative form of the modified M13mp19 was digested with Bam HI and EcoRI. The desired fragments were purified by gel electrophoresis, ligated, and transformed into *E. coli* strain DH5 alpha (BRL). *E. coli* transformed with phage that carried the insert were selected on ampicillin plates. The phage so produced were termed JD32.

25

The plasmid form of the phage, pJD32 (M13mp19Amp<sup>r</sup>), was mutagenized so that two restriction sites, EagI and KpnI, were introduced into gene III without altering the amino acids encoded in this region. The restriction sites were introduced using standard PCR *in vitro* mutagenesis techniques as described by Innis, M., et al. in PCR Protocols—A Guide to Methods and Applications (1990), Academic Press, Inc.

30

The KpnI site was constructed by converting the sequence, TGTTCC, at position 1611 to GGTACC. The two oligonucleotides used to effect the mutagenesis have the following sequence:

LP159: AAACTTCCTCATGAAAAAGTC

LP162: AGAATAGAAAGGTACCACTAAAGGA

To construct the EagI restriction site, the sequence at position 1631 of pJD32, CCGCTG, was changed to CGGCCG using the following two oligonucleotides:

LP160: TTTAGTGGTACCTTCTATTCTCACTCGGCCGAAACTGT  
LP161: AAAGCGCAGTCTCTGAATTTACCG

More specifically, the PCR products obtained using the primers LP 159, LP 162 and LP 160 and LP161 were digested with BspHI and KpnI, and KpnI and AlwNI, respectively. These were ligated with T4 ligase to M13mp19 previously cut with BspHI and AlwNI to yield M13mpLP66. This vector contains the desired EagI and KpnI restriction sites, but lacks the ampicillin resistance gene,  $\beta$ -lactamase. Thus, the vector M13mpLP67, which contains the EagI and KpnI restriction sites and  $\beta$ -lactamase was produced by removing the  $\beta$ -lactamase sequences from pJD32 by digesting the vector with XbaI and EcoRI. The  $\beta$ -lactamase gene was then inserted into the polylinker region of M13mpLP66 which was previously digested with XbaI and EcoRI. Subsequent ligation with T4 ligase produced M13mpLP67, which was used to generate the random peptide library. Figures 1 and 2 schematically sets forth the construction of M13mpLP67.

### Example III

#### Production of Phage Encoding Random Peptides

To produce phage having DNA sequences that encode random peptide sequence, M13LP67 was digested with EagI and KpnI, and ligated to the oligonucleotides produced as described, in Example I, above. The ligation mixture consisted of digested M13LP67 DNA at 45 ng/ $\mu$ l, a 5-fold molar excess of oligonucleotides, 3.6 U/ $\mu$ l of T4 ligase (New England Biolabs), 25 mM Tris, pH 7.8, 10 mM MgCl<sub>2</sub>, 2 mM DTT, 0.4 mM ATP, and 0.1 mg/ml BSA. Prior to being added to the ligation mixture, the individual oligonucleotides were combined and heated to 95°C for 5 minutes, and subsequently cooled to room temperature in 15  $\mu$ l aliquotes. Next, the ligation mixture was incubated for 4 hours at room temperature and subsequently overnight at 15°C. This mixture was then electroporated into *E. coli* as described below.

M13LP67 DNA was electroporated into H249 cells that were prepared essentially as described by Dower, W., J. J. Miller, J. F. and Ragsdale, C. W., 1988, Nucleic Acids Research, 16:6127. H249 cells are a recA, sup<sup>+</sup>, F' kan<sup>R</sup> derivative of MM294. Briefly, 4 x 10<sup>9</sup> H249 cells and 1  $\mu$ g of M13LP67 DNA were combined in 85  $\mu$ l of a low conductivity solution consisting of 1 mM Hepes. The cell/M13LP67DNA mixture was positioned in a chilled 0.56 mm gap electrode of a

BTX electroporation device (BTX Corp.) and subjected to a 5 millisecond pulse of 560 volts.

Immediately following electroporation, the cells were removed from the electrode assembly, mixed with fresh H249 lawn cells, and plated at a density of about 5  $2 \times 10^5$  plaques per 400 cm<sup>2</sup> plate. The next day phage from each plate were eluted with 30 ml of fresh media, PEG precipitated, resuspended in 20% glycerol, and stored frozen at -70°C. About  $2.8 \times 10^7$  plaques were harvested and several hundred analyzed to determine the approximate number that harbor random peptide sequences. Using the polymerase chain reaction to amplify DNA in the region that encodes the random peptide sequence, it was determined that about 50-90 % of the phage contained a 69 base pair insert at the 5' end of gene III. This confirmed the presence of the oligonucleotides that encode the random peptides sequences. The PCR reaction was conducted using standard techniques and with the following oligonucleotides:

5' TCG AAA GCA AGC TGA TAA ACC G 3'

15 5' ACA GAC AGC CCT CAT AGT TAG CG 3'

The reaction was run for 40 cycles after which the products were resolved by electrophoresis in a 2% agarose gel. Based on these results, it was calculated that phage from the  $2.8 \times 10^7$  plaques encode about  $2 \times 10^7$  different random amino acid sequences.

20

#### Example IV

#### Properties of Phage Encoded Random Peptides

The random peptide library was screened for peptide binding activity to streptavidin as follows. Streptavidin was bound to a solid matrix and used to screen for 25 phage that carry streptavidin binding peptides on their surface using biopanning techniques well known in the art. This procedure was essentially conducted as described by Parmley S. F. and Smith, G. P., 1988, *Gene*, 73:305.

Briefly, 60 mm polystyrene plates (60 x 15 mm, Falcon Corp.) were coated overnight with 1mg/ml of streptavidin in 0.1 M NaHCO<sub>3</sub>, pH 8.6, containing 0.02% 30 Na N<sub>3</sub> to prevent bacterial growth. The next day the streptavidin solution was removed, and the plates were blocked for at least one hour with 10 ml of a blocking solution consisting of 29 mg/ml of bovine serum albumin, 3 µg/ml streptavidin, 0.02 % NaN<sub>3</sub>, in 0.1 M NaHCO<sub>3</sub>, pH 8.6. Next, the plates were rinsed with TBS-Tween and 10<sup>12</sup> phage adsorbed onto the plates for 15 minutes, followed by washing the plates 35 ten times with TBS-Tween to remove phage that did not bind to streptavidin. TBS-Tween consisted of 50 mM Tris-HCl, pH 7.5, 150 mM NaCl, and 0.5% Tween 20.

The adherent phage were eluted from the plates with 800  $\mu$ l of a sterile solution consisting of 6 M Urea in 0.1 N HCl. (pH adjusted to 2.2 with glycine) The phage were eluted in this solution for 15 minutes, and the solution neutralized with 23  $\mu$ l of 2 M Tris Base. This procedure yielded about  $4 \times 10^5$  phage. A stock of the phage was prepared by reinfection of *E. coli* and preparation of a plate stock, using standard procedures, and the streptavidin biopanning selection procedure repeated. In the second selection,  $10^{10}$  phage were panned, and  $10^8$  were finally eluted. These phage were plated at low density, and 60 separate phage stocks were prepared from randomly selected, individual plaques.

10        The streptavidin binding properties of the 60 isolates were studied in detail. To insure that the phage did indeed display a peptide sequence with streptavidin binding activity, an experiment was conducted to determine if the isolates could be enriched from an excess of phage that did not carry random peptide inserts. Thus, each of the isolates was mixed with M13mp19, and the mixtures were panned on streptavidin plates, and adherent virus eluted as described above. The ratio of random peptide encoding phage to M13mp19 in the initial mixture before biopanning, and the ratio in the eluate was compared by plating the two phage populations on Xgal plates. The two populations could be distinguished because M13mp19 phage form blue plaques, while M13LP67 random peptide sequence expressing phage form white plaques. It was observed that 56 of the 60 isolates were enriched at least 10 fold over M13mp19. A control was run in which both populations of phage were biopanned, as described above, with the exception that the polystyrene dishes were not treated with streptavidin. No enrichment for the random peptide expressing phage was observed. Further, Table 1 shows that 9 of the 60 isolates are enriched from  $8 \times 10^3$  to  $6.7 \times 10^4$  fold in the presence of M13mp19. The Table also shows that 2 of the isolates failed to bind to streptavidin.

Table 1

		- Biotin			+ Biotin		
		Number of M13mp19 Plaques Per Isolate Plaque			Number of M13mp19 Plaques Per Isolate Plaque		
	Isolate	Initial	Eluate	Enrichment	Initial	Eluate	Enrichment
10	A	28	0.002	$1.4 \times 10^4$			
	B	58	0.003	$1.9 \times 10^4$			
	C	19	0.0007	$2.6 \times 10^4$			
	D	60	0.0009	$6.7 \times 10^4$			
	E	140	0.012	$1.2 \times 10^4$			
15	F	23	0.0091	$2.5 \times 10^3$			
	G	28	0.0009	$3.1 \times 10^4$			
	H	26	0.0024	$1.1 \times 10^4$			
	I	16	0.002	$8.0 \times 10^3$			
20	Y	11	5	2.2			
	Z	15	4	3.8			
	M13LP67	9	5	1.8			

To further characterize the streptavidin binding properties associated with certain of the isolates, the biopanning experiment with M13mp19 was repeated but with the addition of 1  $\mu$ M biotin to the panning solution. Biotin binds very tightly to streptavidin, and may be expected to reduce the binding activity of the selected phage to streptavidin. Indeed, Table 1 shows a large reduction in enrichment for random peptide expressing phage when biotin is present.

The DNA that encodes the random peptide sequences from the isolates A through I, shown in Table 1, was sequenced, and the predicted amino acid sequences of the random sequences is shown in Table 2. It is apparent that they exhibit a histidine-proline consequence sequence. The random peptide inserts of 6 isolates that did not exhibit streptavidin binding were also sequenced, and they did not display the histidine-proline consensus sequence.

Table 2 presents the predicted amino acid sequence of the random peptides encoded by several streptavidin binding phage.

Table 2

Isolate	Frequency			
5	A	3	SDDWWHD	HPQN
	B	1	MLWYSPHSFS	HPQN
	C	1	SWWLSW	HPQN
	D	5	ISFENTWLW	HPQF
	E	1	LC	HPQF
10	F	2	PC	HPQY
	G	2	QPFL	HPQG
	H	1	ALCCLSSP	HPNG
	I	4	LN	HPMD
	Consensus			HP

15 To confirm that the consensus sequence, histidine-proline, is indeed responsible for binding to streptavidin, a peptide was synthesized that has the consensus sequence and tested for inhibition of streptavidin binding as described above. The peptide was observed to significantly inhibit streptavidin binding. The peptide was synthesized using known methods, and had the following sequence:

20 Leu-Asn-His-Pro-Met-Asp-Asn-Arg-Leu-His-Gly-COOH

Example VConstruction of Random Peptide Library in λ Phage

Ten µg of λ gt 11 Sfi-Not was cut with EcoRI and Not 1, phenol/chloroform  
25 extracted, ethanol precipitated, washed in 80% ethanol, and resuspended in 10 µl TE buffer. We then mixed 12.5 picomoles of each of the following two oligonucleotides in 2 µl of water.

5'-GGCCGCTCAATCAGTCAXXXXXXXXXXXXXXXXXXXXXXX  
XXXXXXXXXXXXXXXXXXXXGGCGGCCGGAGGTGGCGG-3'  
30 5'-AATTCTCCGCCACCTCCGCCNNNSNNNSNNNSNNNSNNNN  
SNNSNNSNNSNNNSNNNSTGACTGATTGACG-3'

Next, the oligonucleotides were heated to 80°C for 2 minutes cooled, and mixed with an additional 10 µl of H<sub>2</sub>O, 2 µl of 10 x Ligase mix (described by Maniatis, above), 5 µl of the cut DNA described above and 1 µl of T4 DNA ligase (England  
35 Biolabs, 100,000 - 500,000 U/ml) and incubated at room temperature for 3 hours, and subsequently at 15°C for 2.5 days. Sixty percent of the mixture was then loaded onto a 0.5% agarose TEA gel in one 5 x 5 x 1.5 mm well. After electrophoresis, the high molecular weight DNA band was excised from the gel, and placed in a solution consisting of 100 mM NaCl, 30 mM Tris, pH 8 for 1 to 12 hours. The gel fragment  
40 was then removed from the blotted with a Kim wipe to remove excess solution and

melted at 70°C and 5 µl of the molten agarose was placed in each of 2 tubes. The tubes were heated to 90°C for 2 minutes and 6 minutes, respectively, and incubated at 65°C for 2 hours. 2.5 µl from each tube was packaged using Stratagene Corporation's Giga pack plus phage and packing kit. These packaged phage were used to infect Y1090  
5 cells (Promega) and plated at a density of 50,000 plaques per 150 mm diameter plate. 4 x 10<sup>6</sup> independent plaques were generated by this method and plate stocks were prepared.

The above λ phage library was screened with a monoclonal antibody to feline leukemia virus that recognizes the peptide QAMGPQLVL. The antibody is described  
10 by Nunberg, J. M. et al., 1984, PNAS, 81:3675, while the peptide is described by Kuldip, S. et al., 1989, in Peptides--Chemistry, Structure and Biology, (eds., Jean Rivier and Garland Marshall) Escom, Leiden , 1990. Standard λ gtl1 screening techniques were used, as described by Davis, R. W., and Young, R. A., in U. S.  
15 Patent No. 4,788,135. More than 6 different phage were isolated. The antibody bound to protein on nitrocellulose filters that had been overlaid on plaques created by these phage, thus indicating the presence of a peptide sequence in the phage. This binding could be greatly reduced by co-incubating the filters and antibodies with 25 µg/ml of the peptide QAMGPQLVL, but an irrelevant peptide had no effect.

The present invention has been described with reference to specific  
20 embodiments. However, this application is intended to cover those changes and substitutions which may be made by those skilled in the art without departing from the spirit and the scope of the appended claims.

**WE CLAIM:**

1. A protein(s) comprising a random peptide sequence(s) as part of said protein(s), wherein said random peptide sequence(s) is characterized by not deleteriously altering the biological properties of said protein(s).
2. A protein(s) as described in claim 1 comprising a random peptide sequence(s) as part of said protein(s), wherein said sequence(s) has more than 6 and less than 16 amino acid residues.
- 10 3. A protein(s) as described in claim 2, wherein said random peptide sequence(s) comprise about 15 amino acid residues.
4. A protein(s) as described in claim 3, wherein said random peptide sequence(s) are near the amino terminal end of said protein(s).
- 15 5. A protein(s) as described in claim 4, further comprising at the carboxyl terminal end of the random peptide sequence(s) an effective number of spacer amino acids.
6. A protein(s) as described in claim 5, wherein said spacer amino acids comprise proline.
- 25 7. A protein(s) as described in claim 6, wherein said spacer amino acids comprise about 6 prolines.
8. A viral surface fusion protein(s) comprising a random peptide sequence(s) as part of said protein(s), wherein said protein(s) are encoded by an oligonucleotide comprising the formula:  
30 V----(NNS)<sub>X</sub>----L----V<sub>1</sub>  
wherein V stands for deoxynucleotides that encode one or more amino acid(s) residues found at the amino terminus of a viral surface protein V<sub>1</sub>, X is more than 6 and less than 16 amino acid residues encoded by an equal mixture of A, G, C and T deoxynucleotides, N, and an equal mixture of C and G deoxynucleotides, S, and L  
35 stands for about 6 linker amino acid residues.

9. A nucleic acid sequence(s) that encode(s) protein(s) comprising random peptide sequence(s) as part of said protein(s), wherein said random peptide sequence(s) is characterized by not substantially deleteriously altering the biological properties of said protein(s).

10. A nucleic acid sequence(s) that encode protein(s) comprising random peptide sequences as part of said proteins, wherein said random peptide sequence(s) have more than 6 and less than 16 amino acid residues.

10 11. A nucleic acid sequence(s) as described in claim 10, wherein said nucleic acid sequence(s) encode random peptide sequence(s) comprising about 15 amino acid residues.

12. A nucleic acid sequence(s) as described in claim 11, wherein said  
15 encoded random peptide sequence(s) are near the amino terminal end of said protein(s).

13. A nucleic acid sequence(s) as described in claim 12, further comprising at the carboxyl terminal end of the random peptide sequence(s) an effective number of spacer amino acid residues.

14. A nucleic acid sequence(s) as described in claim 12, wherein said spacer amino acid residues comprise proline.

15. A nucleic acid sequence(s) as described in claim 14, wherein said spacer  
25 amino acid residue comprise about 6 prolines.

16. An oligonucleotide pair, comprising a first and second oligonucleotide, said first oligonucleotide comprising nucleic acid sequence(s) that encode a random peptide sequence, and said second oligonucleotide comprising deoxyinosine nucleic acid sequence(s) that bind to said random peptide nucleic acid sequence(s).

17. A method for producing a random peptide library, comprising the steps  
of:

- 35 a) synthesizing an oligonucleotide pair, said oligonucleotide pair comprising a first and second oligonucleotide, said first oligonucleotide comprising nucleic acid sequence(s) that encode a random peptide sequence, and said second oligonucleotide comprising deoxyinosine

nucleic acid sequence(s) that bind to said random peptide nucleic acid sequence(s);

- b) fusing said oligonucleotide pair to a target nucleic acid sequence(s) that encodes a protein(s) in an expression vector to produce a fusion protein(s) that contains said random peptide sequence(s); and
- c) inserting said expression vector into a host cell capable of expressing said fusion protein(s).

18. A method for producing a random peptide library as described in claim  
10 17, wherein said first oligonucleotide further comprises the formula:



wherein V stands for deoxynucleotides that encode one or more amino acid residue(s) found at the amino terminus of a viral surface protein  $V_1$ , N is an equal mixture of A, G, C and T deoxynucleotides, S is an equal mixture of C and G deoxynucleotides, X corresponds to a number that yields a random peptide sequence that is characterized by not substantially deleteriously altering the biological properties of said protein(s) in said expression vector, and L stands for an effective number of linker amino acids.

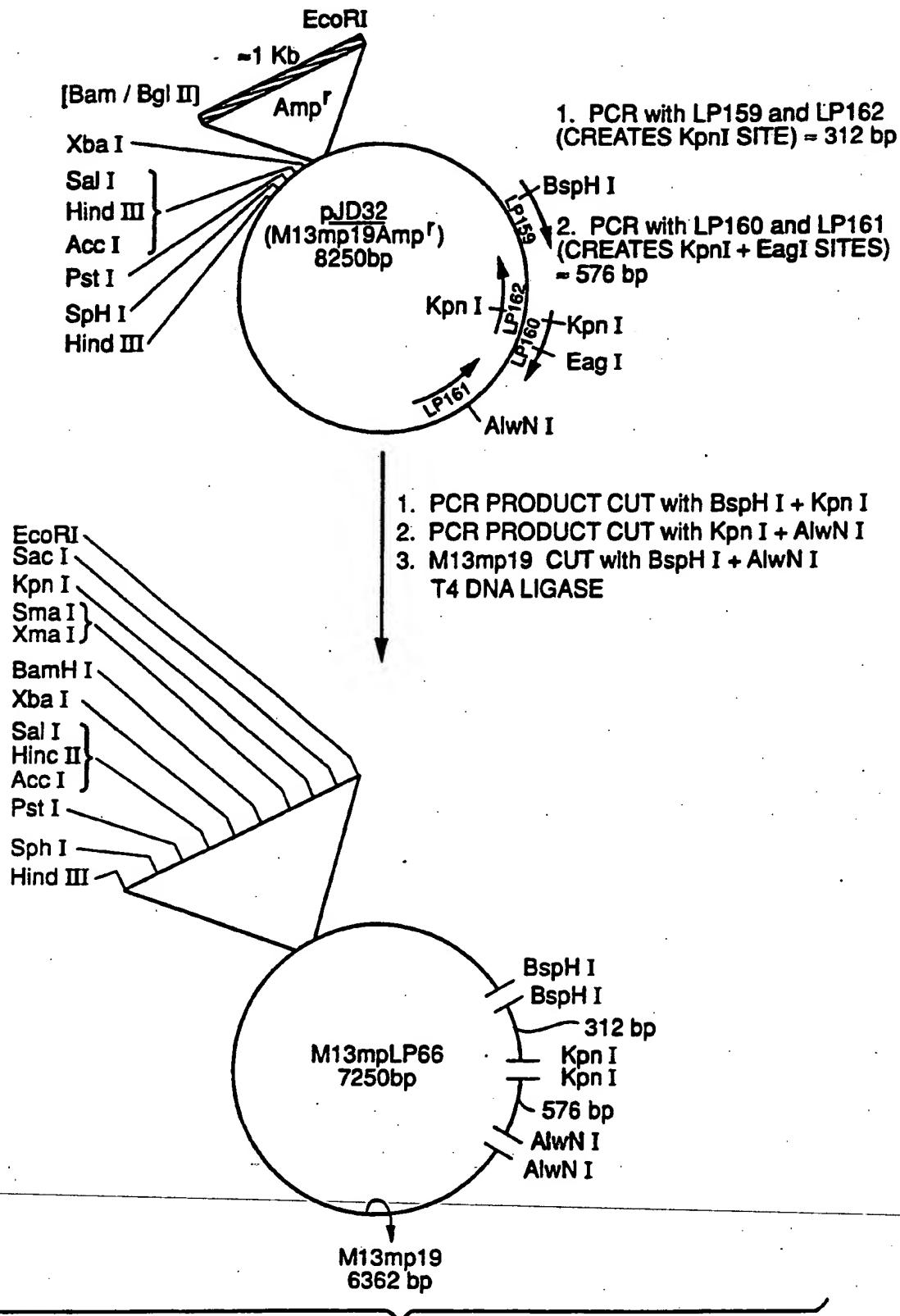
19. A method for producing a random peptide library as described in claim  
20 18, wherein said effective number of linker amino acids is about 6.

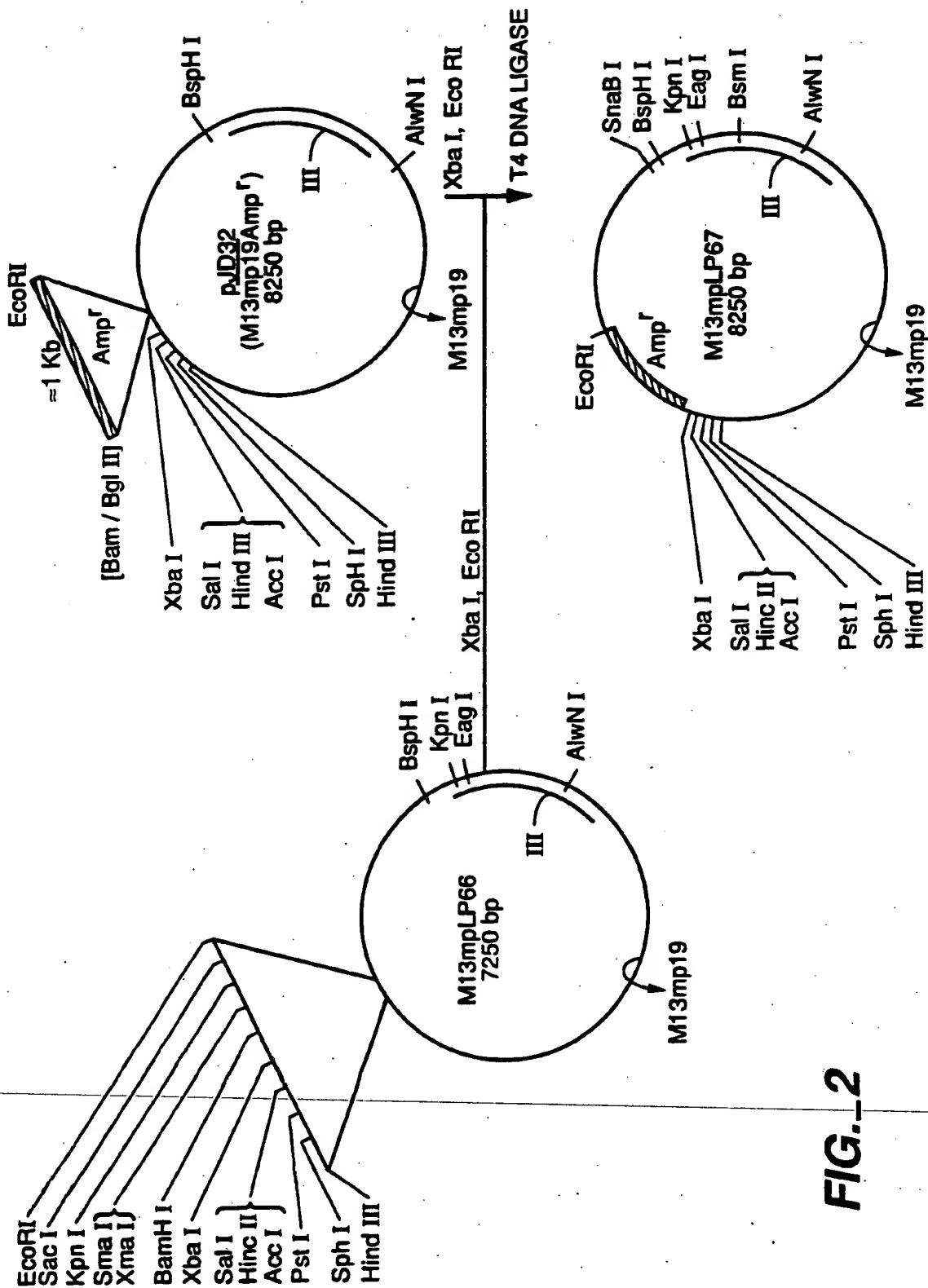
20. A method for producing a random peptide library as described in claim  
19, wherein said linker amino acids are proline.

25 21. A method for producing a random peptide library as described in claim  
20, wherein said viral surface protein is a filamentous viral surface protein(s).

22. The vector M13LP67 comprising a DNA sequence that encodes a random peptide library with ATCC accession number \_\_\_\_\_.

23. A peptide that binds to streptavidin comprising the consensus sequence Histidine-Proline.

**FIG.\_1****SUBSTITUTE SHEET**

**FIG. 2****SUBSTITUTE SHEET**

## INTERNATIONAL SEARCH REPORT

International Application No.

PCT/US 91/03332

I. CLASSIFICATION OF SUBJECT MATTER (If several classification symbols apply, indicate all) <sup>6</sup>

According to International Patent Classification (IPC) or to both National Classification and IPC C 12 N 15/10,  
 IPC<sup>5</sup>: 15/62, 15/34, 15/70, 15/00 // C 07 K 7/06

## II. FIELDS SEARCHED

Minimum Documentation Searched <sup>7</sup>

Classification System	Classification Symbols
IPC <sup>5</sup>	C 12 N

Documentation Searched other than Minimum Documentation  
to the Extent that such Documents are included in the Fields Searched <sup>8</sup>

III. DOCUMENTS CONSIDERED TO BE RELEVANT<sup>9</sup>

Category <sup>10</sup>	Citation of Document, <sup>11</sup> with indication, where appropriate, of the relevant passages <sup>12</sup>	Relevant to Claim No. <sup>13</sup>
P,X	Science, volume 249, 27 July 1990, (Lancaster, US), J.J. Devlin et al.: "Random peptide libraries: A source of specific protein binding molecules", pages 404- 406 see the whole document --	1-23
X	Journal of Biological Chemistry, volume 264, no. 34, 5 December 1989, The American Society for Biochemistry and molecular, Inc., (Baltimore, US), B.D. Lemire et al.: "The mirochondrial targeting function of randomly gene- rated peptide sequences correlates with predicted helical amphiphilicity" pages 20206-20215 see the abstract -- --	1,2,4,9,10 12 . / .

<sup>10</sup> Special categories of cited documents: <sup>10</sup>

- <sup>"A"</sup> document defining the general state of the art which is not considered to be of particular relevance
- <sup>"E"</sup> earlier document but published on or after the international filing date
- <sup>"L"</sup> document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- <sup>"O"</sup> document referring to an oral disclosure, use, exhibition or other means
- <sup>"P"</sup> document published prior to the international filing date but later than the priority date claimed

<sup>11</sup> "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention<sup>12</sup> "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step<sup>13</sup> "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.<sup>14</sup> "Z" document member of the same patent family

## IV. CERTIFICATION

Date of the Actual Completion of the International Search

22nd July 1991

Date of Mailing of this International Search Report

- 9. 09. 91

International Searching Authority

EUROPEAN PATENT OFFICE

Signature of Authorized Officer

M. PEIS

M. Peis

## III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

Category *	Citation of Document, ** with indication, where appropriate, of the relevant passages	Relevant to Claim No.
Y	<p>Gene, volume 73, 1988, Elsevier Science Publishers, B.V., (Amsterdam, NL), S.F. Parmley et al.: "Antibody-selectable filamentous fd phage vectors: purification of target genes", pages 305-318      see summary; page 316, left-hand column, line 45 - right-hand column, line 13      cited in the application</p> <p>--</p>	8
Y	<p>Protein Engineering, volume 2, no. 4, 1988, (Eynsham, GB), I.S. Dunn et al.: "Improved peptide function from random mutagenesis over short 'windows'", pages 283-291      see abstract; page 284, left-hand column, lines 1-15</p>	8
X	--	1,2,9,10
A	<p>The Journal of Histochemistry and Cytochemistry, volume 34, no. 11, 1986, The Histochemical Society, Inc, (N.Y., US) L. Scopsi et al.: "Nonspecific immuno-cytochemical reactions with certain neurohormonal peptides and basic peptide sequences", pages 1469-1475      see page 1471, left-hand column, line 50 - page 1472, left-hand column, line 13.</p> <p>--</p>	23
X	<p>Science, volume 24, 1988, (Lancaster, US), J.F. Reidhaar-Olson et al.: "Combinatorial cassette mutagenesis as a probe of the informational content of protein sequences", pages 53-57      see figure 2      cited in the application</p> <p>--</p>	16
A	<p>US, A, 4713366 (VERNON C. STEVENS)      15 December 1987      see claim 2</p> <p>-----</p>	5-7

ANNEX TO THE INTERNATIONAL SEARCH REPORT  
ON INTERNATIONAL PATENT APPLICATION NO. US 9103332  
SA. 48374

This annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report.  
The members are as contained in the European Patent Office EDP file on 27/08/91.  
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US-A- 4713366	15-12-87	US-A- 4855285	08-08-89